International Information and
Engineering Technology Association
Advancing the World of Information and Engineering

Modified Ant Colony Optimization for Human Recognition in Videos of Low Resolution

Katakam Ranganarayana[1*], Gurrala Venkateswara Rao[2]

[1] Department of IT, GITAM Deemed to be University, Visakhapatnam 530045, India
[2] Department of CSE, GITAM Deemed to be University, Visakhapatnam 530045, India

Corresponding Author Email: katakam916@gmail.com

ABSTRACT

Privacy protections for people filmed in public settings is a prerequisite to widespread camera use. For this reason, low-resolution videos are used from which specific people can be reliably obscured. Since the human region in low-resolution videos comprises of so few pixels and so little information, human detection is more challenging there than it is in high-resolution videos. With the current state of affairs, one of the most important challenges is tracking a target from lower resolution movies. Identification or monitoring of persons in low-resolution movies has become a common issue in many domains due to a lack of appropriate data. This study presents a novel people-detection algorithm that makes use of low-resolution film to overcome the aforementioned problem. In the first stage, a three-step procedure is executed, the video data gathered from low-resolution videos from various form of data is considered. The captured video is separated into frames and transformed from RGB to gray-scale. Local Binary Pattern (LBP) method is used in the second phase to accomplish background subtraction. Thirdly the feature extraction is performed in which histogram of optical flow (HOF) and some of the features are extracted in the form of eigen values. Finally these features are optimized using Modified Ant Colony Optimization (MACO) model to remove the unwanted features and select global features. Finally, classification operation is performed using Support Vector Machine (SVM) classifier to recognize the person from lower resolution videos. The results obtained using the implemented MACO-SVM obtains good results when compared with existing techniques with rate of accuracy 91.46% for soccer dataset, 90.8% for KTH dataset and 89.75% accuracy using VIRAT dataset.

1. INTRODUCTION

Many efforts have been done in the field of identifying the objects in motion and compliance over the last few decades to make the following dependable, robust, and successful: surveillance of videos, robotics, multifactor authentication, multimedia creation, biomedical sciences, and so on [1]. However, there are other problems that present roadblocks to enhancing these applications. As analyzed by Jiang et al. [2] these issues might include lighting changes, dynamic backgrounds, concealment, space closure, shadow, and so on. These obstacles grow more severe while tracking of objects in videos with lower resolution. It's tough to identify exactly what you're looking for in a low-resolution video since a majority of distinguishing data, such as visual and unique content is lost. It results to the incorrect investigation, which leads to the discovery of a defective occurrence. Some of the major benefits by using low resolution videos for performing research is been discussed by Park et al. [3] and they are less memory used for storage of videos, time for transmission of data will be reduced and lesser time interval. The majority of typical tracking algorithms rely on higher resolution video (HRV) to extract a straight line as designed by Chen et al. [4] and Cremers et al. [5] the bag of words. However, because they function in higher-resolution frames, these approaches necessitate additional computation expenses. Other approaches in the literature employ lesser quality movies as

input, but these videos are eventually enhanced to greater resolution with the use of higher resolution techniques, demonstrating that they are less expensive. Many approaches in atypical human detection literature, such as employ categories of classification for recognizing events rather than lower-resolution input video [6, 7]. The training database needs to closely watch and the researchers need study time. Other solutions, analyzed by Lili Cui et al. [8], need human configuration at the start of the event's default program and have substantial computing expenditures. According to the preceding literature, a novel comprehensive algorithm for detecting objects in low quality videos is being developed, which will aid in the development of a completely automated surveillance system.

The author introduced two effective methods for recognition of movement in lower resolution films recorded from data under surveillance implemented by Ranganarayana et al. [9], with the goal of providing security. However, a set of limitations are induced for every algorithm. The Standard FCM method, for example, does not incorporate any sort of information which is spatial in the context of image, causing the data to be vulnerable to noise and other imaging abnormalities. Second, the mean shift method is only reliant on colour features, limiting its effectiveness to background subtraction. Further SVM classification is used by Schuldt et al. [10] for identification of human in low resolution videos. The evolution of optimization techniques helps in improving

the detection rate by selecting the global best features. Ranganarayana et al. [11] proposed particle swarm optimization techniques for identification of human in lower resolution videos.

Some worries concerning data privacy are also raised. An ever-increasing variety of cameras, from those installed in security and protection systems to those in wearable gadgets and even in our own mobile phones, are constantly recording footage in both public and private spaces. It's bad enough that people are recording video, but then putting it away in the cloud. For reasons of security, it is not advisable to keep or post such films to distant servers. To solve this problem, videos might be sent at the lowest quality necessary for identification or analysis. The difficulty of efficient activity recognition with extremely low-resolution frames arises from the inability of existing algorithms to adapt well to these constraints due to substantial changes in extracted characteristics.

In this paper, a modified any colony optimization technique is implemented to extract the global best features and classified using SVM in which the training and testing of data will be performed. Further in section 2, the process of proposed MACO is demonstrated clearly with the work process. In section 3, the datasets used are been discussed briefly with number of samples. In section 4, the process of proposed framework is designed and in section 5 the results are discussed.

2. MODIFIED ANT COLONY OPTIMIZATION

To achieve its goals, ACO relies on two mechanisms: trail evaporation and the possible intervention of daemons. The first mechanism aids in lowering all trail values over time to limit the unwarranted growth of trails among the other components. The second mechanism is used to carry out coordinated tasks that single ants cannot, such as trying to launch a local optimization strategy or upgrading global evidence to determine either to bias the strategic plan of research from a non-local perspective [12] or moving to next process. To be more specific, an ant is a fundamental computational agent that generates a solution tailored to the problem at hand in order to solve it iteratively.

ACO uses a group of software agents referred to as artificial ants to look for optimal solutions to an optimization problem. The optimization model is recast as a difficulty of detecting the optimal route across a weighted graph in order to use ACO. Graph movement is used by the synthetic ants to create solutions gradually. Graph components have run-time values adjusted by the ants, hence the solution creation process is unpredictable and biassed by a pheromone model.

Partially solved problems are referred to as states. At the heart of the ACO algorithm is a loop in which each ant travels (performs a step) from one state ι to another ψ, equivalent to a more comprehensive temporary result. The step function is denoted as σ, every ant is termed as k and the computed set is a feasible extension of its present position which is termed as $A_k^\sigma(l)$, these terms move to some of the probabilities. The distribution function involved in this process will have probabilities and are as follows. For ant 'k', the probability function is given as $p_{l\psi}^k$, i.e. ant moving from one state ι to the other state ψ, this moving character is the result of a combination of two factors.

- One is the *attractiveness* $\eta_{\iota\psi}$ of the move, as determined by some criteria reflecting the attractiveness of that action;
- Other one is the *trail level* $\tau_{\iota\psi}$ of the move representing how proficient it was in the previous to accomplish that precise move: as a result, it gives a probability indicator of the action's attractiveness.

After all the ants have reached their solution the updating of trails are performed, with the level of trails increasing or decreasing to correlate to moves that were partial progress of "good" or "poor" solutions, respectively. The writers working on the ACO method have detailed the broad structure just described in various ways.

The algorithm of ANTS is as follows, based on the factors mentioned.

step1. Compute a (linear) lower bound LB to the problem and initialize $\tau_{l\psi}(\forall \iota, \psi)$ with the primal variable values

step2. For $k = 1, min$ which $m = number\ of\ ants$

evaluate $\eta_{l\psi}(\forall \iota, \psi)$

identify the probability of the move

record the move chosen by ant to the k-th ant's tabu list

The recorded move is resembles till the solution is achieved by ant k

obtained solution should be to its local optimum

end for

step3. For every movement of ant from one state to other state $(\iota\psi)$,

evaluate $\Delta\tau_{l\psi}$ and update trails by means of $\Delta\tau_{l\psi} = a.\tau_0.\left(1 - \frac{z_{curr}-LB}{z-LB}\right)$

where, a is random number of features previously selected, $'z'$ is the average of the last solution of $'k'$ and LB is a lower limit on the cost of the ideal issue solution.

step4. If not satisfied (end_test) **switch to** step 2.

3. DATASETS USED FOR IMPLEMENTATION

3.1 VIRAT

VIRAT video dataset is one of the most important datasets utilised in the improvement of the vision-based community [9]. The dataset is supposed to be highly realistic and hard for processing of video based domains, as it contains low-resolution films and will conduct person recognition in video tasks. For doing experimental assessments, the Virat dataset provides video sequences ranging in duration from 0.5sec to 5min. In this research, huge number of video sequences which is around 5000 and each having a frame rate of 30 frames per second. The frame has a width of 1280 pixels and a height of 720 pixels.

3.2 KTH

The KTH database is the prominent and widely used data set for signals categorization of individual activity [10]. The database comprises actions such as walking, running, and jogging that are conducted on a variety of themes and in a variety of contexts. The dataset includes 600 films in each category, six distinct acts, and four different circumstances. With a vertical camera at 25 frames per second, each sequence serves as a distinct background (frames per second). In this

dataset the video sequences available are of shot range having duration of 5seconds. The frame has a width of 160 pixels and a height of 120 pixels.

3.3 SOCCER

Kicking, running, strolling, and dribbling are the four activities in the soccer dataset. There are 255 videos accessible in this collection. The video clip is shortened to 3 seconds in duration and 28 frames per second. The frame has a width of 150 pixels and a height of 150 pixels.

4. PROPOSED METHODOLOGY

The implementation structure of proposed methodology is shown in Figure 1. Every step shown in Figure 1 is step wise approach to identify the human from low resolution videos. The methodology includes optimization algorithm and machine learning technique. The entire process is evaluated by using the datasets which are publicly available as discussed in section 3.
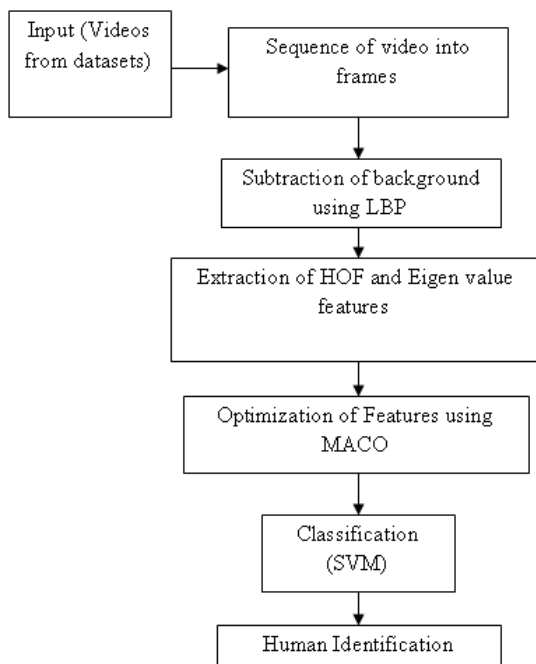

Figure 1. Flow diagram of the work

4.1 Formation of frames

The videos from the discussed datasets are considered as inputs. Each input video has 'N number of frames. The initial and first step to process the data is to extract the frames from the video. These are frames are considered for the process of research. Depending on the frames the recognition of required task is achieved.

4.2 Subtraction of Background

Each and every frame has different background elements or components. These background components reduce the visual perception in identifying the required task. Hence background subtraction is performed before extracting the features. The technique used for modeling the background of every frame is local binary pattern (LBP). The divides frames of a video in

which the first frame is utilized as the background model in the BS process, and we compute LBP for each pixel in the frame. The LBP of the following frame is compared to the LBP of the backdrop model. The background model LBPs is comparable with any of the value, the pixels in the frame are classed as background; otherwise the pixel in the frame is classified as foreground.

4.3 Feature extraction

The features from the background subtracted frames need to be extracted. The information which is in form of dynamic are accessible by obtaining HOF features form the motion range video frames. In histogram of optical flow the vector fields are accessed which are extremely important in recognizing spatial motions of pictures over time and providing crucial information which is necessary for the work. These HOF feature extracted is illustrated in the study [13].

The features extracted are evaluated by calculating the eigen values of the features. These eigen values calculations depend on the block number per frame and the number of frames for each volume. The changes sensed in the pixels of a certain direction must be established before the calculation of Eigen values. The Eigen vector is the direction having the highest change in Eigen values. As a consequence, these Eigen values and Eigen vectors are produced for all frames in the video in all directions. These characteristics help in the identification of the individual.

4.4 Optimized features

In this step optimized features which are useful for recognition of human are derived. The technique used to achieve the optimized features is Modified Ant Colony Optimization (MACO). The extracted features are processed using MACO to obtain the global best feature set. The features are selected randomly using modified ACO. The process of feature selection using MACO is done using the algorithm which is discussed in section 2.

4.5 Classification technique

The characteristics generated using the optimization process is classified as training and testing data. In the last stage, the employment of machine learning techniques aids in the improvement of the recognition system. In this article, SVM classifier is suggested for classifying and identification of human. The hyper-plane is used by SVM classifiers to accomplish their operations. Following that, the SVM makes a choice by supplying a test dataset. The parameters are chosen using K-fold cross-validation. A non-linear SVM with RBF kernel is suggested and the dataset is divided into 70% for training and 30% for testing. As a result, person identification is carried out. Based on the functioning of SVM the performance metrics of the implemented technique is evaluated.

5. RESULTS AND DISCUSSION

The suggested technique is evaluated with the help of three different datasets they are KTH, Soccer and VIRAT. The common point of all three datasets is every dataset has low resolution videos. Hence these benchmark datasets can be used to evaluate the suggested techniques in identification of human

from the videos. These datasets have different video length and with a specific time length. The parameters are evaluated using these datasets to prove the efficiency of the implemented work. The below figures show the recognition results obtained used proposed methodology. In Figure 2, the results using soccer dataset is seen. In Figure 3 and Figure 4 are the output results obtained using KTH and VIRAT datasets.
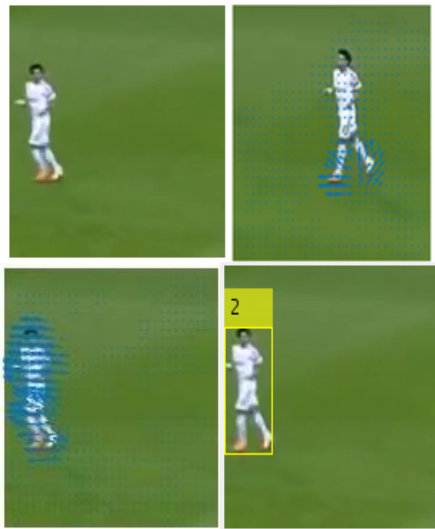

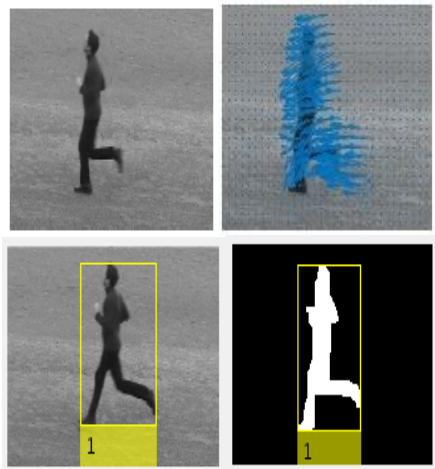**Figure 2.** Recognition of person using soccer dataset

Figure 3. Recognition of person using KTH dataset

Figure 4. Recognition of person using VIRAT dataset

5.1 Performance Metrics to prove the efficiency of proposed work

The evaluation metrics are utilized to differentiate between the previous and developed techniques. The parameters that are shown in Table 1 is used to evaluate the improvement of implemented technique. For evaluating different parameters the dataset which is considered need to be divided into training and testing phases. SVM is modeled to recognize the human from videos. The parameters calculated shows the working efficiency of the model which is designed. By evaluating the parameters of the proposed work on three datasets helps in proving the ability of technique in tracing out the humans.

i. Miss Rate (MR): The miss rate is given by number of false positives in positive samples by total number of positive samples.

MR= 1-Recall

ii. FPPI (False Positive per Image): The log-average Miss Rate is comparable to the MAP you stated in that it pertains to the items that are not identified.

iii. Detection Rate: Number of correctly classified person recognition

iv. Precision: Fraction of true positives /All positives

v. False Detection Rate:1-Precison

vi. F1 Measure: TP/(TP+ (0.5*FP+FN))

The results obtained using SVM classifier without the interversion of optimization technique is shown in Table 1. The results obtained using modified ant colony optimization are tabulated and shown on Table 2.

Table 1. Evaluation of datasets without using support vector machine classifier

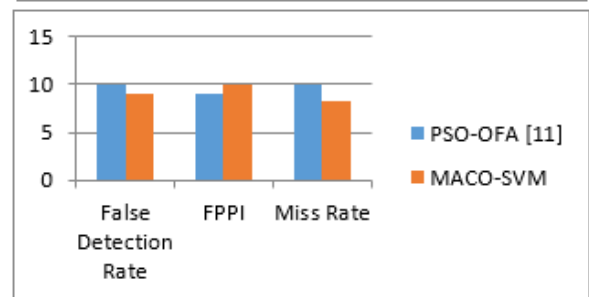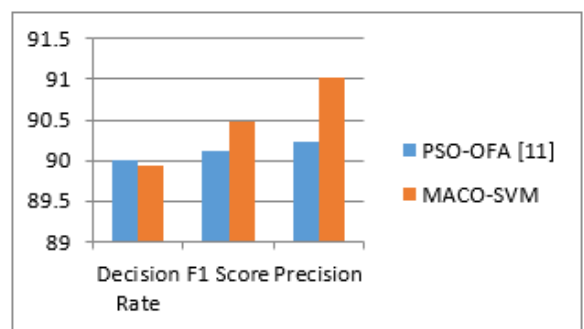|	KTH	Soccer	VIRAT
Accuracy	85.29	84.15	82.63
Rate of detection	83.78	82.61	81.67
False Detection rate	14.37	15.56	18.16
FPPI	13.28	14.39	16.48
Miss Rate	16.21	17.38	18.32
F1 Score	84.69	83.514	81.754
Precision	85.628	84.431	81.836

Figure 5. Comparison results obtained for KTH dataset

The use of optimization technique along with SVM classifier aids in the improvement of the rate of accuracy and other evaluation metrics for the recognition of human in lower quality videos. The achieved parameter values are provided in Table 2 and compared to the previous swarm optimization approach.
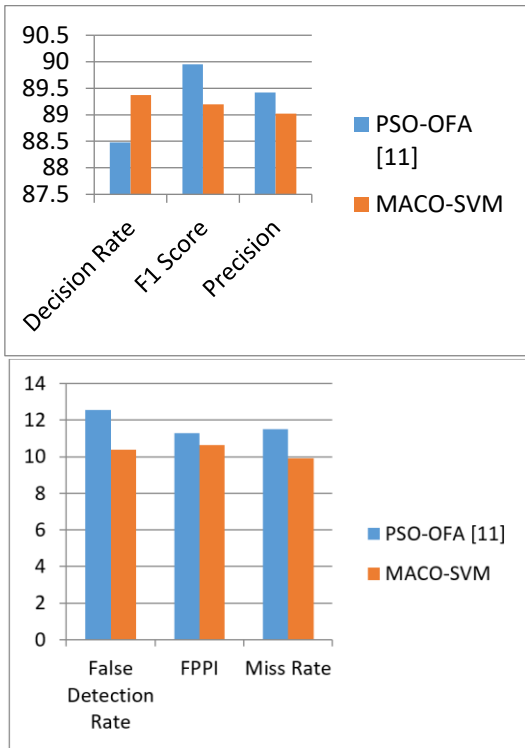
represented in Table 3.

The graphical representation of accuracy results obtained in recognition of human from videos using different methods and different public available datasets is shown in Figure 8. The proposed optimization with machine learning technique provide better rate of accuracy i.e. 91.46%.
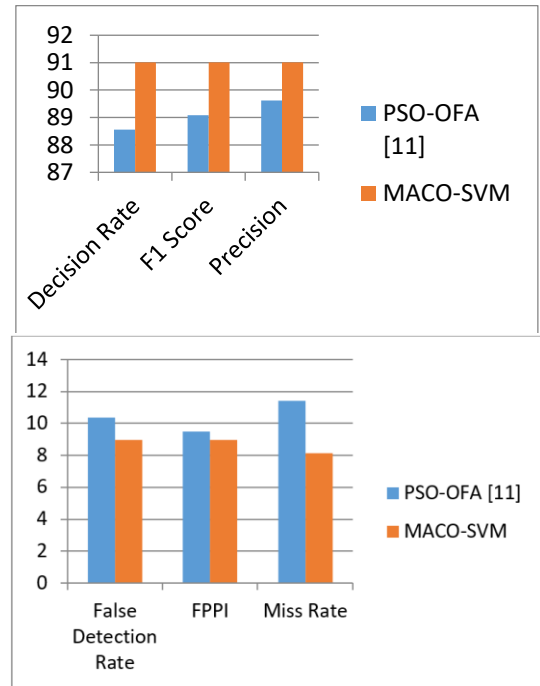
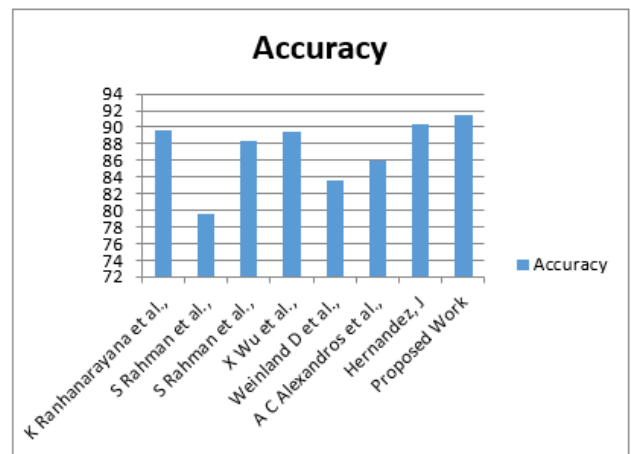**Figure 6.** Comparison results obtained for Virat dataset

Figure 7. Comparison results obtained for Soccer dataset

The comparison of the results obtained are shown using visual representation in the form of bars graphs. Here, in Figure 5 the accuracy, decision rate, F1 score, Precison are compared as one group and False detection rate, FPPI and miss rate comparison in another group.

Figure 6 shows that when compared to PSO-SVM, implemented MACO-SVM procedures perform well in terms of all assessment measures. The rate of accuracy acquired for the VIRAT dataset using the MACO approach is 89.75, which is 7% higher than the value obtained without utilising the optimization technique and also 1.1% higher when compared with swarm optimization [11]. The results obtained using soccer dataset for proposed MACO is compared with existing swarm technique and is shown in Figure 7. The comparison of accuracy results obtained using different techniques is

Figure 8. Accuracy levels

Table 2. Performance analysis of implemented methodology

		KTH	Soccer	VIRAT
Rate of Detection	PSO-SVM [11]	90.019	88.56	88.48
	MACO-SVM	**89.94**	**91.01**	**89.37**
False Detection rate	PSO-SVM [11]	9.98	10.37	12.57
	MACO-SVM	**8.98**	**8.98**	**10.97**
FPPI	PSO-SVM [11]	9.04	9.50	11.27
	MACO-SVM	**10.04**	**8.98**	**10.62**
Miss Rate	PSO-SVM [11]	9.98	11.43	11.51
	MACO-SVM	**8.22**	**8.13**	**9.90**
F1 Score	PSO-SVM [11]	90.12	89.08	89.95
	MACO-SVM	**90.47**	**91.01**	**89.20**
Precision	PSO-SVM [11]	90.23	89.62	89.42
	MACO-SVM	**91.01**	**91.01**	**89.02**

Table 3. Comparison of accuracy results obtained using different techniques

Accuracy (%)	PSO-SVM [11]	89.56
	HOG+LBP+BOW [14]	79.63
	HOF+LBP+FV [14]	88.43
	spatio-temporal context distribution [15]	89.4
	HOG descriptors [16]	83.5
	Silhoutte based Key poses [17]	85.9
	Feature Extraction based on local descriptors [18]	90.3
	Proposed MACO-SVM	**91.46**

6. CONCLUSION

This study looks at the challenge of recognizing people in low-resolution movies. The fundamental problem of person recognition has been enhanced by leveraging HOF, Eigen values, and modified Ant colony optimization approaches. The feature extraction approach is critical in recognizing the individual human. The optimization strategies aid in acquiring the greatest characteristics on a global scale. The suggested technique is tested using datasets from KTH, VIRAT, and Soccer. The correctness of a study work can be checked by using different datasets. In terms of accuracy and other characteristics, the suggested MACO-SVM approach outperforms. More characteristics can be combined with the current ones to improve recognition performance.

REFERENCES

[1] Srinivasan, K., Pokumaran, K., Sainarayan, G. (2009). Improved background subtraction techniques for security in video application, Anti-counterfeiting, Security, and Identification in Communication, pp. 114-117. https://doi.org/10.1109/ICASID.2009.5276945

[2] Jiang, N., Liu, W.Y., Su, H., Wu, Y. (2011). Tracking low resolution objects by metric preservation, Computer Visionand Pattern Recognition (CVPR), pp. 1329-1336. http://dx.doi.org/10.1109/CVPR.2011.5995537

[3] Park, H.S., Shi, J. (2015). Social saliency prediction. In Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Boston, MA), pp. 4777-4785.

[4] Chen, Y., Rui, Y., Huang, T. (2006). Multicue HMM-UKF for real time contour tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1525 - 1529. https://doi.org/10.1109/tpami.2006.190

[5] Cremers, D. (2006). Dynamical statistical shape priors for level set-based tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1262-1273. https://doi.org/10.1109/tpami.2006.161

[6] Kamijo, S., Ikeuchi, K., Sakauchi, M. (2000). Traffic monitoring and accident detection at intersections. IEEE Transactions on Intelligent Transportation Systems, pp. 108-118. https://doi.org/10.1109/6979.880968

[7] Wang, T., Snoussi, H. (2012). Histograms of optical flow orientation for visual abnormal events detection. IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance (AVSS), pp. 13-18. https://doi.org/10.1109/AVSS.2012.39

[8] Cui, L.L., Li, K.H., Chen, J.P., Li, Z.B. (2011). Abnormal event detection in traffic video surveillance based on local features. Image and Signal Processing (CISP), pp. 362-366. https://doi.org/10.1109/CISP.2011.6099933

[9] Ranganarayana, K. and Venkateswara Rao, G. (2019). Motion detection in low resolution video surveillance data to provide personal privacy. IJAER, 14(23): 4251-4255.

[10] Schuldt, C., Laptev, I., Caputo, B., (2004). Recognizing human actions: A local SVM approach. ICPR'04. https://doi.org/10.1109/ICPR.2004.1334462

[11] Ranganarayana, K., VenkateswaraRao, G. (2021). Human recognition using 'PSO-OFA' in low resolution videos. Turkish Journal of Computer and Mathematics Education, 12(11): 697-703.

[12] Dorigo, M., Stützle, T. (2002) The ant colony optimization metaheuristic: Algorithms, applications and advances. Handbook of Metaheuristics, Kluwer Academic Publishers, Norwell, MA, pp 251-285.

[13] Zhou, D.X., Zhang, H. (2005). Modified GMM background modeling and optical flow for detection of moving objects. Conference Proceedings of IEEE International Conference on Systems, Man and Cybernetics, 3: 2224-2229. https://doi.org/10.1109/ICSMC.2005.1571479

[14] Rahman, S., See, J. Ho, C. (2017). Exploiting textures for better action recognition in low quality videos. J Image Video Proc., 2017(1). https://jivp-eurasipjournals.springeropen.com/articles/10.1186/s13640-017-0221-2.

[15] Wu, X., Xu, D., Duan, L., Luo, J. (2011). Action recognition using context an appearance distribution features. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 489-496. http://dx.doi.org/10.1109/CVPR.2011.5995624

[16] Weinland, D., Ozuysal, M., Fua, P. (2010). Making action recognition robust to occlusions and viewpoint changes. In: Daniilidis, K., Maragos, P., Paragios, N. (Eds.), Computer Vision ECCV 2010, Lecture Notes in Computer Science, Springer, Berlin/Heidelberg, 635-648.

[17] Chaaraoui, A.A., Climent-Pérez, P. Flórez-Revuelta, F. (2013). Silhouette-based human action recognition using sequences of key poses. Pattern Recognition Letters, 34(15): 1799-1807. https://doi.org/10.1016/j.patrec.2013.01.021

[18] Hernandez, J., Montemayor, A., Pantrigo, J., Sanchez, A. (2011). Human action recognition based on tracking features. In: Ferrandez, J., Alvarez Sanchez, J., de la Paz, F., Toledo, F. (Eds.), Foundations on Natural and Artificial Computation, Lecture Notes in Computer Science, 6686, pp. 471-480.